

**Original citation:**

Ge, Tian, Luo, Qiang, Grabenhorst, Fabian, Feng, Jianfeng and Rolls, Edmund T..  
(2013) Attention-dependent modulation of cortical taste circuits Revealed by granger  
causality with signal-dependent noise. PLoS Computational Biology, Volume 9 (number  
10). Article number e1003265. ISSN 1553-7358

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/57892>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the  
University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 (CC BY 3.0)  
license and may be reused according to the conditions of the license. For more details  
see: <http://creativecommons.org/licenses/by/3.0/>

**A note on versions:**

The version presented in WRAP is the published version, or, version of record, and may  
be cited as it appears here.

For more information, please contact the WRAP Team at: [publications@warwick.ac.uk](mailto:publications@warwick.ac.uk)



<http://wrap.warwick.ac.uk>

# Attention-Dependent Modulation of Cortical Taste Circuits Revealed by Granger Causality with Signal-Dependent Noise

Qiang Luo<sup>1,2</sup>, Tian Ge<sup>3,4</sup>, Fabian Grabenhorst<sup>5</sup>, Jianfeng Feng<sup>1,3,4,\*</sup>, Edmund T. Rolls<sup>4,6,\*</sup>

**1** Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, PR China, **2** College of Information Systems and Management, National University of Defense Technology, Hunan, PR China, **3** Centre for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai, PR China, **4** Department of Computer Science, University of Warwick, Coventry, United Kingdom, **5** Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom, **6** Oxford Centre for Computational Neuroscience, Oxford, United Kingdom

## Abstract

We show, for the first time, that in cortical areas, for example the insular, orbitofrontal, and lateral prefrontal cortex, there is signal-dependent noise in the fMRI blood-oxygen level dependent (BOLD) time series, with the variance of the noise increasing approximately linearly with the square of the signal. Classical Granger causal models are based on autoregressive models with time invariant covariance structure, and thus do not take this signal-dependent noise into account. To address this limitation, here we describe a Granger causal model with signal-dependent noise, and a novel, likelihood ratio test for causal inferences. We apply this approach to the data from an fMRI study to investigate the source of the top-down attentional control of taste intensity and taste pleasantness processing. The Granger causality with signal-dependent noise analysis reveals effects not identified by classical Granger causal analysis. In particular, there is a top-down effect from the posterior lateral prefrontal cortex to the insular taste cortex during attention to intensity but not to pleasantness, and there is a top-down effect from the anterior and posterior lateral prefrontal cortex to the orbitofrontal cortex during attention to pleasantness but not to intensity. In addition, there is stronger forward effective connectivity from the insular taste cortex to the orbitofrontal cortex during attention to pleasantness than during attention to intensity. These findings indicate the importance of explicitly modeling signal-dependent noise in functional neuroimaging, and reveal some of the processes involved in a biased activation theory of selective attention.

**Citation:** Luo Q, Ge T, Grabenhorst F, Feng J, Rolls ET (2013) Attention-Dependent Modulation of Cortical Taste Circuits Revealed by Granger Causality with Signal-Dependent Noise. PLoS Comput Biol 9(10): e1003265. doi:10.1371/journal.pcbi.1003265

**Editor:** Olaf Sporns, Indiana University, United States of America

**Received:** January 19, 2013; **Accepted:** August 20, 2013; **Published:** October 24, 2013

**Copyright:** © 2013 Luo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** QL is partly supported by grants from the National Natural Sciences Foundation of China (No. 11101429, No. 11271121, No. 71171195), Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20114307120019), and National Basic Research Program of China (No. 2011CB707802). FG was supported by the Gottlieb-Daimler- and Karl Benz-Foundation. JF is a Royal Society Wolfson Research Merit Award holder, partially supported by the National Centre for Mathematics and Interdisciplinary Sciences (NCMIS) of the Chinese Academy of Sciences and Key Program of National Natural Science Foundation of China (No. 91230201). The fMRI investigation was supported by the McDonnell Centre for Cognitive Neuroscience at the University of Oxford, and was performed at the Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) at the University of Oxford. Support was also received from the Oxford Centre for Computational Neuroscience. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: Jianfeng.Feng@warwick.ac.uk (JF); Edmund.Rolls@warwick.ac.uk (ETR)

## Introduction

In the past decade, Granger causality (GC) has emerged as a widely used method for causal inferences, and has been applied to biological time series obtained from many different types of investigation, for example, to the fMRI blood-oxygen level dependent (BOLD) signals to detect effective connectivity between brain areas and thus to shed light on how the brain works [1–4]. The basic idea of GC can be traced back to Wiener [5], who conceived the notion that if the prediction of one time series can be improved by incorporating the past history of a second one, then the second time series has a causal influence on the first. Granger later formulated this idea in the context of linear autoregressive (AR) models [6]. GC is completely data-driven and based on time precedence. The interactions discovered by GC may be unidirectional or reciprocal. GC is easy to implement, relies on a small set of straightforward assumptions, and does not need any knowledge about how the data are generated. Therefore,

it can be applied directly to almost any time series data [7]. However, over-simplification of the model may result in an incorrect use or interpretation of GC and even spurious causal inferences in some situations [8–10]. Care is therefore needed in the use of GC.

One possible over-simplification in some scenarios is that the covariance matrix of the noise, conditional on the past history of the time series and the noise process, is assumed to be time invariant. For example, spike trains of neurons are typically close to Poisson processes in their timing, and the variance thus increases linearly with the signal [11,12]. Similar conditionally heteroskedastic data have been observed in many physiological recordings, such as the data collected from patients with epilepsy and Parkinson's disease [13]. Therefore, it is natural to conjecture that changes in the volatility of one time series may have an impact on the mean activity or volatility of another time series, which indicates that causal influences may be evident in the second order statistics. Clearly, these causal relationships cannot be captured by

## Author Summary

We show that in cortical areas such as the insular, orbitofrontal, and lateral prefrontal cortex, the variation of the blood-oxygen level dependent (BOLD) time series across trials measured with functional magnetic resonance imaging (fMRI) increases with the magnitude of the signal. We describe a new method of measuring causal effects with Granger causality that takes into account this signal-dependent noise. We show in a functional neuroimaging investigation with the new method that there is a causal influence from the anterior lateral prefrontal cortex that during attention to the pleasantness of taste stimuli increases the response of the orbitofrontal cortex to the taste; and there is a causal influence from the posterior lateral prefrontal cortex to the insular taste cortex during attention to the intensity of taste stimuli. This shows how part of the circuitry involved in the effects of selective attention on the pleasantness and intensity of stimuli operates in the brain.

classical GC based on a simple AR model, which does not deal with time series data with changing volatility (variance). Moreover, although it has been widely observed and investigated that the signal-dependent noise plays important roles in neuronal activities [14–16], it is still unclear whether this property carries through to fMRI BOLD signals, after the neuronal signals are delayed and smoothed by the haemodynamic response function.

In this paper, we provide empirical evidence that the variance of the noise in the fMRI BOLD time series increases linearly with the square of the signal in a number of cortical areas, such as the insular taste, orbitofrontal, and lateral prefrontal cortical areas. In this context we present a Granger causal model with signal-dependent noise to detect GC in both the mean and variance of data with time varying volatility. We also propose a likelihood ratio test to infer GC with signal-dependent noise accurately and efficiently. We show, by simulation studies, that this novel method substantially outperforms classical GC when signal-dependent noise is present.

The new method is evaluated with an fMRI investigation [17] to identify the source of the top-down selective attentional control that differentially biases brain systems involved in affective vs sensory analysis [17–19]. Instructions to pay attention to and later rate the pleasantness of a taste increased the activations to taste stimuli measured with fMRI in the orbitofrontal and pregenual cingulate cortices [17], where the subjective pleasantness of taste is represented [20–24], but not the primary taste cortex in the anterior insula [17], where the subjective intensity and identity of taste are represented [20–22,24–26]. Instructions to pay attention to and later rate the intensity of a taste increased the activations to taste in the insular taste cortex but not in the orbitofrontal and pregenual cingulate cortices [17]. Our new method reveals how the effective top-down connectivity changes when attention is paid to the pleasantness vs the intensity of a taste, and helps in the interpretation of the source of the signals that implement top-down attention.

## Materials and Methods

### Granger causality with signal-dependent noise

**Classical Granger causality.** We start with a brief review of classical GC. Consider the following zero-mean vector autoregressive model (VAR) of order  $p$ :

$$\mathbf{x}_t = \sum_{j=1}^p \mathbf{A}_{x,j} \mathbf{x}_{t-j} + \boldsymbol{\epsilon}_{x,t}, \quad t=1,2,\dots, \quad (1)$$

where  $\mathbf{x}_t$  is a  $d_x$ -dimensional column random vector,  $\mathbf{A}_{x,j}$ ,  $j=1,\dots,p$  are fixed  $d_x \times d_x$  coefficient matrices, and  $\boldsymbol{\epsilon}_{x,t}$  is a  $d_x$ -dimensional independent identically distributed (i.i.d) white noise or innovation process, with a positive definite and time invariant covariance matrix  $\boldsymbol{\Sigma}_x$ . We require that this VAR( $p$ ) process is stable, that is,

$$\det(\mathbf{I} - \mathbf{A}_{x,1}z - \dots - \mathbf{A}_{x,p}z^p) \neq 0, \quad |z| \leq 1, \quad (2)$$

where  $\det(\cdot)$  is the determinant of a matrix,  $\mathbf{I}$  is an identity matrix, and  $z$  is a complex variable. This stability condition implies that the VAR( $p$ ) process is weakly stationary, i.e., its first and second order moments exist and are time invariant [27].

Now, assume  $\mathbf{x}_t$  and  $\mathbf{y}_t$  admit a jointly stable VAR representation.  $\mathbf{x}_t$  can thus be modeled as

$$\mathbf{x}_t = \sum_{j=1}^p \mathbf{A}_{xx,j} \mathbf{x}_{t-j} + \sum_{j=1}^p \mathbf{A}_{xy,j} \mathbf{y}_{t-j} + \boldsymbol{\epsilon}_{xy,t}, \quad t=1,2,\dots, \quad (3)$$

where  $\mathbf{y}_t$  is a  $d_y$ -dimensional column random vector, and  $\boldsymbol{\epsilon}_{xy,t}$  is a white noise process with a covariance matrix  $\boldsymbol{\Sigma}_{xy}$ .

Classical GC depends on temporal precedence and predictability. The idea is that a cause cannot come after the effect. Thus, if  $\mathbf{y}_t$  affects  $\mathbf{x}_t$ , including the past information of  $\mathbf{y}_t$  should improve the predictions of  $\mathbf{x}_t$ . More formally, if the prediction error of  $\mathbf{x}_t$  is reduced when the past information of  $\mathbf{y}_t$  is taken into account, then  $\mathbf{y}_t$  has a causal influence on  $\mathbf{x}_t$  in the sense of Granger. Formulating the idea in the context of a VAR model, the causal influence from  $\mathbf{y}_t$  to  $\mathbf{x}_t$  in the time domain can be quantified as [28,29]

$$\mathcal{F}_{y \rightarrow x} = \log \left[ \frac{\det(\boldsymbol{\Sigma}_x)}{\det(\boldsymbol{\Sigma}_{xy})} \right]. \quad (4)$$

$\mathcal{F}_{y \rightarrow x} > 0$  indicates a causal influence from  $\mathbf{y}_t$  to  $\mathbf{x}_t$ , and  $\mathcal{F}_{y \rightarrow x} = 0$  otherwise. Note that model (1) is a restricted version of model (3), and that  $\mathbf{y}_t$  does not cause  $\mathbf{x}_t$  if and only if  $\mathbf{A}_{xy,j} \equiv \mathbf{0}$  for all  $j=1,\dots,p$  [27].

When the white noise is Gaussian distributed, it has been shown that the GC measure in Eq. (4) is equivalent to the likelihood ratio test statistic [30]

$$\mathcal{R}_{y \rightarrow x} = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{restricted}} | \{\mathbf{x}_t\}_{t=1}^T)}{\mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T)}, \quad (5)$$

where  $\mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{restricted}} | \{\mathbf{x}_t\}_{t=1}^T) \equiv \mathbf{Pr}(\{\mathbf{x}_t\}_{t=1}^T | \hat{\boldsymbol{\theta}}_{\text{restricted}})$  is the likelihood function, i.e., the probability of the observed time series  $\{\mathbf{x}_t\}_{t=1}^T$ , given the maximum likelihood estimate of the parameters  $\hat{\boldsymbol{\theta}}_{\text{restricted}}$  of the restricted model (1).  $\mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T) \equiv \mathbf{Pr}(\{\mathbf{x}_t\}_{t=1}^T | \hat{\boldsymbol{\theta}}_{\text{full}}, \{\mathbf{y}_t\}_{t=1}^T)$  is interpreted similarly under the full model (3). Therefore, a likelihood ratio test can be used for causal inference:

$$r_{y \rightarrow x} = -2 \left[ \log \mathcal{L}(\hat{\theta}_{\text{restricted}} | \{x_t\}_{t=1}^T) - \log \mathcal{L}(\hat{\theta}_{\text{full}} | \{x_t\}_{t=1}^T, \{y_t\}_{t=1}^T) \right]. \quad (6)$$

The test statistic  $r_{y \rightarrow x}$  is approximately chi-squared distributed, with degrees of freedom  $df_{\text{full}} - df_{\text{restricted}}$ , where  $df_{\text{full}}$  and  $df_{\text{restricted}}$  are the number of free parameters of the full model (3) and the restricted model (1), respectively.

**The signal-dependent noise model.** To relax the assumption of a time invariant covariance matrix in the AR model, Engle invented the first changing volatility model — the autoregressive conditional heteroskedasticity (ARCH) model [31], which was then extended to generalized ARCH (GARCH) models [32,33] as well as multivariate cases [34–36]. Assume  $\mathbf{r}_t$  is a  $d$ -dimensional zero mean, serially uncorrelated process, which may be the residual process of some dynamic model and can be represented as

$$\mathbf{r}_t = \Sigma_{t|t-1}^{1/2} \boldsymbol{\epsilon}_t, \quad (7)$$

where  $\boldsymbol{\epsilon}_t$  is a  $d$ -dimensional i.i.d white noise process,  $\Sigma_{t|t-1}$  is the conditional covariance matrix of  $\mathbf{r}_t$ , given  $\Omega_{t-1} = \{\mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \dots\}$ , and  $\Sigma_{t|t-1}^{1/2}$  is the symmetric positive definite square root of  $\Sigma_{t|t-1}$ . Then a multivariate ARCH process of order  $q$  takes the form

$$\text{vech}(\Sigma_{t|t-1}) = \gamma_0 + \Gamma_1 \text{vech}(\mathbf{r}_{t-1} \mathbf{r}_{t-1}^\top) + \dots + \Gamma_q \text{vech}(\mathbf{r}_{t-q} \mathbf{r}_{t-q}^\top), \quad (8)$$

where  $\text{vech}(\cdot)$  denotes the half-vectorization operator which stacks the columns of a square matrix from the diagonal downwards in a vector,  $\gamma_0$  is a  $\frac{1}{2}d(d+1)$ -dimensional column vector of constants and  $\Gamma_j, j=1, \dots, q$  are  $\frac{1}{2}d(d+1) \times \frac{1}{2}d(d+1)$  coefficient matrices. It can be seen that even for a bivariate series with a low order, this general model has a fairly large number of parameters. Therefore, more restricted models were proposed. For example, Bollerslev et al. considered diagonal ARCH processes where all the  $\Gamma_j$  matrices are diagonal [35]. To guarantee the positive definiteness of the conditional covariance matrix  $\Sigma_{t|t-1}$ , Baba, Engle, Kraft and Kroner investigated the following variant of a multivariate ARCH model, known as the BEKK model [37,38]

$$\Sigma_{t|t-1} = \Gamma_0 + \Gamma_1^\top \mathbf{r}_{t-1} \mathbf{r}_{t-1}^\top \Gamma_1 + \dots + \Gamma_q^\top \mathbf{r}_{t-q} \mathbf{r}_{t-q}^\top \Gamma_q, \quad (9)$$

where  $^\top$  is the matrix transpose,  $\Gamma_0 = C_0^\top C_0$  is positive definite, and all  $\Gamma_j, j=1, \dots, q$  are  $d \times d$  matrices. In contrast to the diagonal model, the BEKK model produces interactions between second order moments and can generate rich volatility dynamics.

We now present a Granger causal model with signal-dependent noise [13]. Consider the following multivariate model with time varying volatility, in particular, signal-dependent noise:

$$\begin{aligned} \mathbf{x}_t &= \sum_{j=1}^p \mathbf{A}_{x,j} \mathbf{x}_{t-j} + \mathbf{r}_{x,t}, & \mathbf{r}_{x,t} &= \Sigma_{x,t|t-1}^{1/2} \boldsymbol{\epsilon}_{x,t}, \\ \Sigma_{x,t|t-1} &= \mathbf{C}_x^\top \mathbf{C}_x + \sum_{j=1}^q \mathbf{B}_{x,j}^\top \mathbf{x}_{t-j} \mathbf{x}_{t-j}^\top \mathbf{B}_{x,j}, \end{aligned} \quad (10)$$

where  $\mathbf{x}_t$  is a  $d_x$ -dimensional column random vector,  $\boldsymbol{\epsilon}_{x,t}$  is a  $d_x$ -dimensional Gaussian distributed white noise process with zero mean and unit variance,  $p$  and  $q$  are the model orders,  $\mathbf{A}_{x,j}, j=1, \dots, p$ ,  $\mathbf{B}_{x,j}, j=1, \dots, q$  and  $\mathbf{C}_x$  are coefficient matrices. The volatility model is a modification of the BEKK model

[38] in which the conditional covariance matrix  $\Sigma_{x,t|t-1}$  does not regress on the residual process  $\mathbf{r}_{x,t}$  but only depends on the past history of the process  $\mathbf{x}_t$  before time  $t$ . Hence, the covariance (second order statistics) of the noise process is coupled to the mean (first order statistics). This form also guarantees the positive definiteness of  $\Sigma_{x,t|t-1}$ . Clearly, when  $\mathbf{B}_{x,j} \equiv \mathbf{0}$  for all  $j$ , the conditional covariance is time invariant and the model reduces to the AR model. In the light of these points, we term our model (10) the AR-BEKK model.

We now summarize our use of the terms ‘signal’ and ‘noise’ in the reminder of the paper for the model and in the empirical analysis of the fMRI data. We assume that the observed time series  $x_t$  is a realization from the following general process:

$$x_t = f(\mathbf{U}_{t-1}) + \sigma^{1/2}(\mathbf{U}_{t-1}) \epsilon_t, \quad (11)$$

where  $\mathbf{U}_t = \{x_t, x_{t-1}, \dots\}$ ,  $\epsilon_t$  is an i.i.d white noise process,  $f(\cdot)$  and  $\sigma(\cdot)$  can be any continuous functions,  $\sigma(\cdot)$  is nonnegative. We define  $f(\mathbf{U}_{t-1})$  to be the ‘signal’, and  $\sigma^{1/2}(\mathbf{U}_{t-1}) \epsilon_t$  to be the ‘noise’. The variance of the noise is  $\sigma(\mathbf{U}_{t-1}) = \text{var}(x_t | \mathbf{U}_{t-1}) = \text{var}(x_t - f(\mathbf{U}_{t-1}))$ . Given the past history of the time series, the signal is a deterministic process while the noise is what cannot be predicted and produces the variation across realizations. Empirically, since  $f(\mathbf{U}_{t-1}) = \mathbf{E}(x_t | \mathbf{U}_{t-1})$ , the signal is estimated by projecting  $x_t$  onto the subspace spanned by  $\mathbf{U}_{t-1}$ . The noise is estimated by the residual of the projection. In the Results section, we investigate different subspaces spanned by  $\mathbf{U}_{t-1}$  to provide empirical evidence for the signal-dependent noise in fMRI BOLD time series. In the model, we specify particular forms of the functions  $f(\cdot)$  and  $\sigma(\cdot)$ . In particular, according to model (10), we assume  $f(\mathbf{U}_{t-1})$  is a linear function of  $x_{t-1}$ , i.e.,  $f(\mathbf{U}_{t-1}) = ax_{t-1}$ , and we assume  $\sigma(\mathbf{U}_{t-1})$  is a quadratic function of  $x_{t-1}$ , i.e.,  $\sigma(\mathbf{U}_{t-1}) = c^2 + b^2 x_{t-1}^2$ . Therefore, in the model, the signal is estimated by  $ax_{t-1}$  and the variance of the noise is estimated by  $c^2 + b^2 x_{t-1}^2$ . In the Results section, we show the concordance of the definitions of ‘signal’ and ‘noise’ in the model and in the empirical data analysis, that is, in spite of the simplified forms of  $f(\cdot)$  and  $\sigma(\cdot)$ , our model captures a large portion of the variance in the empirical signal and noise. We note that classical Granger causality assumes that the variance of the noise as just defined is constant across the time course of the process (e.g., the time course of an fMRI trial), and that Granger causality with signal-dependent noise allows causality to be calculated more powerfully if the variance of the noise within a trial is not constant.

**Granger causality with signal-dependent noise.** To define the causal relationship between  $\mathbf{x}_t$  and another  $d_y$ -dimensional time series  $\mathbf{y}_t$ , consider the following joint AR-BEKK model:

$$\mathbf{z}_t = \sum_{j=1}^p \mathbf{A}_j \mathbf{z}_{t-j} + \mathbf{r}_t, \quad (12)$$

where  $\mathbf{z}_t = (\mathbf{x}_t^\top, \mathbf{y}_t^\top)^\top$ ,  $\mathbf{r}_t = (\mathbf{r}_{x,t}^\top, \mathbf{r}_{y,t}^\top)^\top$ ,  $\mathbf{r}_{xy,t} = \Sigma_{xy,t|t-1}^{1/2} \boldsymbol{\epsilon}_{xy,t}$ ,  $\mathbf{r}_{yx,t} = \Sigma_{yx,t|t-1}^{1/2} \boldsymbol{\epsilon}_{yx,t}$ ,  $\boldsymbol{\epsilon}_{yx,t}$  and  $\boldsymbol{\epsilon}_{xy,t}$  are  $d_x$ -dimensional and  $d_y$ -dimensional independent Gaussian distributed white noise processes with zero mean and unit variance respectively, and

$$\begin{aligned} \Sigma_{xy,t|t-1} &= \mathbf{C}_{xy}^\top \mathbf{C}_{xy} + \sum_{j=1}^q \mathbf{B}_{xy,j}^\top \mathbf{z}_{t-j} \mathbf{z}_{t-j}^\top \mathbf{B}_{xy,j}, \\ \Sigma_{yx,t|t-1} &= \mathbf{C}_{yx}^\top \mathbf{C}_{yx} + \sum_{j=1}^q \mathbf{B}_{yx,j}^\top \mathbf{z}_{t-j} \mathbf{z}_{t-j}^\top \mathbf{B}_{yx,j}. \end{aligned} \quad (13)$$

Here,  $\mathbf{A}_j, j=1, \dots, p$ ,  $\mathbf{B}_{xy,j}, \mathbf{B}_{yx,j}, j=1, \dots, q$  and  $\mathbf{C}_{xy}, \mathbf{C}_{yx}$  are

all coefficient matrices. The causal influence from  $\mathbf{y}_t$  to  $\mathbf{x}_t$  can be defined as [13]

$$\mathcal{F}_{y \rightarrow x} = \log \left[ \frac{\det(\mathbf{C}_x^\top \mathbf{C}_x)}{\det(\mathbf{C}_{xy}^\top \mathbf{C}_{xy})} \right]. \quad (14)$$

$\mathcal{F}_{y \rightarrow x} > 0$  if  $\mathbf{y}_t$  has a causal effect on  $\mathbf{x}_t$ , and  $\mathcal{F}_{y \rightarrow x} = 0$  otherwise. It can be seen that  $\mathbf{y}_t$  can help improve the prediction of  $\mathbf{x}_t$  by impacting on either its mean activity through the coefficients in  $\mathbf{A}_j$ , or its variance through the coefficients in  $\mathbf{B}_{xy,j}$ . These two cases correspond to causality in the mean and variance respectively. When the noise in  $\mathbf{x}_t$  is not signal-dependent, i.e.,  $\mathbf{B}_{xy,j} \equiv 0$  and  $\mathbf{B}_{yx,j} \equiv 0$  for all  $j$ , the model reduces to a VAR model and the definition of causality coincides with classical GC in the time domain (See Eq. (4)).

**Stability conditions.** To guarantee the stability of the AR-BEKK model, we provide the stability condition for a simple first-order model, i.e.,  $p=q=1$  in Eq. (12). This is the model that we usually use for fMRI data analysis considering the poor temporal resolution of BOLD signals, and the relatively fast signal transmission between groups of neurons [2,4,39,40]. In the remainder of this paper, both in simulations and real data analysis, we focus on this first-order model unless otherwise specified. We also assume that  $\epsilon_{xy,t}$  and  $\epsilon_{yx,t}$  are uncorrelated. The stability of the model involves both the first and second order stability conditions, i.e., the unconditional mean and covariance of  $\mathbf{z}_t$  exist and are time invariant. For the first order stability condition, it follows from the theory of the AR model that all the eigenvalues of  $\mathbf{A}_1$  have modulus less than 1. For the second order stability, note that

$$\text{cov}(\mathbf{z}_t | \mathbf{U}_{t-1}) = \text{cov}(\mathbf{r}_t | \mathbf{U}_{t-1}) = \text{diag}\{\boldsymbol{\Sigma}_{xy,t|t-1}, \boldsymbol{\Sigma}_{yx,t|t-1}\}, \quad (15)$$

and

$$\text{cov}(\mathbf{z}_t | \mathbf{U}_{t-1}) = \mathbf{E}(\mathbf{z}_t \mathbf{z}_t^\top | \mathbf{U}_{t-1}) - \mathbf{E}(\mathbf{z}_t | \mathbf{U}_{t-1}) \mathbf{E}(\mathbf{z}_t^\top | \mathbf{U}_{t-1}), \quad (16)$$

where  $\mathbf{U}_{t-1} = (\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots)$ . Therefore,

$$\begin{aligned} \mathbf{E}(\mathbf{z}_t \mathbf{z}_t^\top | \mathbf{U}_{t-1}) &= \\ &\mathbf{A}_1 \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \mathbf{A}_1^\top + \text{diag}\{\mathbf{C}_{xy}^\top \mathbf{C}_{xy} + \mathbf{B}_{xy,1}^\top \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \mathbf{B}_{xy,1}, \\ &\quad \mathbf{C}_{yx}^\top \mathbf{C}_{yx} + \mathbf{B}_{yx,1}^\top \mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top \mathbf{B}_{yx,1}\}. \end{aligned} \quad (17)$$

Taking the expectation on both sides yields

$$\begin{aligned} \mathbf{E}(\mathbf{z}_t \mathbf{z}_t^\top) &= \\ &\mathbf{A}_1 \mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top) \mathbf{A}_1^\top + \text{diag}\{\mathbf{C}_{xy}^\top \mathbf{C}_{xy} + \mathbf{B}_{xy,1}^\top \mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top) \mathbf{B}_{xy,1}, \\ &\quad \mathbf{C}_{yx}^\top \mathbf{C}_{yx} + \mathbf{B}_{yx,1}^\top \mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top) \mathbf{B}_{yx,1}\}. \end{aligned} \quad (18)$$

This can be transformed into the following equation using the vectorization operator, which stacks the columns of a square matrix into a column vector:

$$\begin{aligned} \text{vec}[\mathbf{E}(\mathbf{z}_t \mathbf{z}_t^\top)] &= \\ &\mathbf{A}_1 \otimes \mathbf{A}_1 \text{vec}[\mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top)] + \text{vec}\left[\text{diag}\left\{\mathbf{C}_{xy}^\top \mathbf{C}_{xy}, \mathbf{C}_{yx}^\top \mathbf{C}_{yx}\right\}\right] \\ &+ \left[\text{diag}\left\{\mathbf{B}_{xy,1}^\top, \mathbf{B}_{yx,1}^\top\right\} \otimes \text{diag}\left\{\mathbf{B}_{xy,1}^\top, \mathbf{B}_{yx,1}^\top\right\}\right] \cdot \mathbf{T} \cdot \text{vec}[\mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top)] \\ &:= (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}) \text{vec}[\mathbf{E}(\mathbf{z}_{t-1} \mathbf{z}_{t-1}^\top)] + \tilde{\mathbf{C}}, \end{aligned}$$

where  $\otimes$  is the Kronecker product,

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}.$$

Therefore, it is required that all the eigenvalues of  $\tilde{\mathbf{A}} + \tilde{\mathbf{B}}$  have modulus less than 1.

**Model estimation.** Using Bayes' theorem, the joint density function of  $\mathbf{r}_1, \dots, \mathbf{r}_T$  is

$$f(\mathbf{r}_1, \dots, \mathbf{r}_T) = f(\mathbf{r}_1) f(\mathbf{r}_2 | \mathbf{r}_1) \dots f(\mathbf{r}_T | \mathbf{r}_{T-1}, \dots, \mathbf{r}_1). \quad (19)$$

Thus, the conditional distribution of  $\mathbf{r}_t$  given  $\mathbf{U}_{t-1}$  is Gaussian and if the  $\mathbf{u}_t$  are observed quantities, the log-likelihood function of the AR-BEKK model described by Eq. (12), for a sample  $\mathbf{u}_1, \dots, \mathbf{u}_T$  is given by

$$\begin{aligned} \log \mathcal{L}(\theta_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T) &= \\ \sum_{t=1}^T \log \mathcal{L}_t(\theta_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T), \end{aligned} \quad (20)$$

where  $\theta_{\text{full}}$  is a vector of all unknown parameters of the model (12) and

$$\begin{aligned} \log \mathcal{L}_t(\theta_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T) &= \\ = -\frac{dx+dy}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}_{xy,t|t-1}| - \frac{1}{2} \mathbf{u}_t^\top \boldsymbol{\Sigma}_{xy,t|t-1}^{-1} \mathbf{u}_t, \end{aligned} \quad (21)$$

where the required initial values for specifying  $\boldsymbol{\Sigma}_{t|t-1}$  are assumed to be available. The likelihood function may be maximized with respect to the parameters  $\theta_{\text{full}}$  by using numerical methods. Specifically, the initial values of  $\mathbf{A}_1$  are given by the least square estimates of  $\mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \mathbf{r}_t$ , assuming a simple AR model, and  $\mathbf{B}_{xy,1}, \mathbf{B}_{yx,1}$  and  $\mathbf{C}_{xy}, \mathbf{C}_{yx}$  are then initialized to diagonal matrices whose  $l$ -th element on the diagonal is the least square estimate of

$$[\mathbf{r}_{xy,t}^2]_l = [\mathbf{C}_{xy}^2]_{ll} + [\mathbf{B}_{xy,1}^2]_{ll} [\mathbf{z}_{t-1}^2]_l, \quad (22)$$

and

$$[\mathbf{r}_{yx,t}^2]_l = [\mathbf{C}_{yx}^2]_{ll} + [\mathbf{B}_{yx,1}^2]_{ll} [\mathbf{z}_{t-1}^2]_l, \quad (23)$$

using the residuals  $\mathbf{r}_t$  from the AR fitting. The constrained maximum likelihood estimation of the model parameters can be obtained by solving the optimization problem

$$\hat{\theta}_{\text{full}} = \arg\max_{\theta_{\text{full}}} \sum_{t=1}^T \log \mathcal{L}_t(\theta_{\text{full}}), \quad (24)$$

while satisfying the first and second order stability conditions derived above. We use Matlab function `fmincon` with the interior-point algorithm to tackle this restricted optimization problem. (Note that sometimes `fmincon` with the interior-point algorithm may fail to converge to a reasonable solution. In this case, we use the active-set algorithm as an alternative.) The parameters of the restricted model (10) can be estimated similarly.

**Causal inferences.** The nonparametric bootstrap method for causal inferences used in [13] is time-consuming. Here we develop an analog of the likelihood ratio test for classical GC (See Eq. (6)) to improve computational efficiency. Similarly, the likelihood ratio test statistic takes the form

$$r_{y \rightarrow x} = -2 \left[ \log \mathcal{L}(\hat{\theta}_{\text{restricted}} | \{\mathbf{x}_t\}_{t=1}^T) - \log \mathcal{L}(\hat{\theta}_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T) \right] \quad (25)$$

$$= -2 \sum_{t=1}^T \left[ \log \mathcal{L}(\hat{\theta}_{\text{restricted}} | \{\mathbf{x}_t\}_{t=1}^T) - \log \mathcal{L}(\hat{\theta}_{\text{full}} | \{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{y}_t\}_{t=1}^T) \right].$$

The test statistic approximately follows a chi-squared distribution, and the degrees of freedom are  $2d_x d_y$ . Therefore, a parametric chi-squared test can be carried out to test the significance of the causal influence. This likelihood ratio test also has a connection to the transfer entropy between time series [30]. However, since the residual process in the AR-BEKK model is not a Gaussian white noise, the likelihood ratio test is not equivalent to the measure defined in Eq. (14).

To test the difference between the causalities in the opposite directions between brain areas, note that the difference of the two causality measures is  $r_{\text{diff}} = \frac{1}{2} r_{y \rightarrow x} - \frac{1}{2} r_{x \rightarrow y}$ , where  $r_{y \rightarrow x}$  and  $r_{x \rightarrow y}$  are two chi-squared distributed random variables with the same degrees of freedom. Therefore, the distribution function of  $r_{\text{diff}}$  is

$$T_m(x) = \frac{1}{2^m \sqrt{\pi} \Gamma(m + \frac{1}{2})} x^m K_m(x), \quad (26)$$

where  $K_m$  is a modified Bessel function,  $\Gamma(\cdot)$  is a Gamma function, and  $m = d_x d_y - \frac{1}{2}$  [41]. A table for the two-sided one and five percent quantile of this distribution can be found in [41]. For example, in the investigation of a pair of univariate time series, i.e.,  $d_x = d_y = 1$ , a difference measure of 4.61 implies a  $p$ -value of 0.01.

## Simulation studies

**Methodology assessment.** We illustrate the Granger causal model with signal-dependent noise and the likelihood ratio test by a simulation study. Consider the following first-order AR-BEKK model for two univariate time series  $x_t$  and  $y_t$ :

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1\sqrt{2} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} r_{xy,t} \\ r_{yx,t} \end{bmatrix}, \quad \begin{bmatrix} r_{xy,t} \\ r_{yx,t} \end{bmatrix} = \begin{bmatrix} \Sigma_{xy,t|t-1}^{1/2} \epsilon_{xy,t} \\ \Sigma_{yx,t|t-1}^{1/2} \epsilon_{yx,t} \end{bmatrix},$$

$$\begin{aligned} \Sigma_{xy,t|t-1}^{1/2} &= 1 + [q_1, (1-q_1)q_3][x_{t-1}, y_{t-1}][x_{t-1}, y_{t-1}]^T [q_1, (1-q_1)q_3]^T, \\ \Sigma_{yx,t|t-1}^{1/2} &= 1 + [(1-q_2)q_4, q_2][x_{t-1}, y_{t-1}][x_{t-1}, y_{t-1}]^T [(1-q_2)q_4, q_2]^T, \end{aligned}$$

where  $q_1$  and  $q_2$  are random numbers uniformly distributed in  $[0, 1]$ , and  $q_3$  and  $q_4$  are random numbers with the probability of 0.6 to be 0 and 0.4 to be 1. It is clear that there is a causal influence from  $y_t$  to  $x_t$  if and only if  $(1-q_1)q_3 \neq 0$ , and from  $x_t$  to  $y_t$  if and only if  $(1-q_2)q_4 \neq 0$ .

We generated 100 models with different  $q_i$ ,  $i = 1, \dots, 4$ , and for each model we generated time series of 1000 points with 2 replicates. We then fitted both the classical Granger causal model and the signal-dependent noise model to the data. Using different  $p$ -value thresholds, the performance of the two models was compared by the ROC (Receiver Operating Characteristic) curve [42].

**An illustration of how signal-dependent noise may arise in BOLD activations.** In the Results section below, we provide empirical evidence for the presence of signal-dependent noise in fMRI BOLD time series. In order to illustrate how signal-dependent noise may arise in BOLD activations from the underlying neuronal firing, we performed the following simulations. These simulations were on a simple model developed for the purposes of illustration. The concept is to investigate how the close to Poisson firing of neuronal spikes of neurons in the cortex for a given mean firing rate [11,12] might be reflected in a signal produced by feeding the spiking neuronal activity into a widely used generative biophysical model describing the hemodynamic response [43]. This hemodynamic model links neuronal activity to blood flow and incorporates the well established Balloon model [44,45]. For the Poisson spiking, the variance of the spike counts in a time window increases linearly with (and is equal to) the mean spike count.

We simulated spike trains of neurons following Poisson processes with mean firing rates of 5 Hz, 40 Hz and 80 Hz respectively for 1 second. The spike trains were then fed into the following biophysical model describing the hemodynamic response induced by the neuronal activity [43]:

$$\begin{cases} \frac{dz}{dt} = z - \kappa s - \gamma(f - 1) \\ \frac{df}{dt} = s \\ \tau \frac{dv}{dt} = f - v^{1/\alpha} \\ \tau \frac{dq}{dt} = \frac{\ell}{\rho} \left[ 1 - (1 - \rho)^{1/f} \right] - q v^{1/\alpha - 1} \end{cases} \quad (27)$$

where  $z$  is the input spike trains;  $s$  is the vasodilatory signal;  $f$  is the cerebral blood inflow (CBF);  $v$  is the cerebral blood volume (CBV);  $q$  is the deoxyhemoglobin (dHb) content;  $1/\kappa$  is time constant of signal decay;  $1/\gamma$  is the time constant of the feedback autoregulatory mechanism;  $\tau$  is the mean transit time in the post-capillary venous compartment;  $\alpha$  is the Grubb's parameter and  $\rho$  is the resting net oxygen extraction fraction by the capillary bed. Finally, the observed BOLD time series is a nonlinear function of the CBV and dHb content:

$$x_t = V_0 [k_1(1-q) + k_2(1-q/v) + k_3(1-v)] + \zeta_t, \quad (28)$$

where  $\zeta_t$  is the observation noise,  $V_0 \approx 0.02$  is the resting blood volume fraction,  $k_1 \approx 7\rho$ ,  $k_2 \approx 2$ ,  $k_3 \approx 2\rho - 0.2$  for 1.5-T scanners. All biophysical parameters were set to their typical values ( $\kappa = 0.65$ ;  $\gamma = 0.41$ ;  $\tau = 0.98$ ;  $\alpha = 0.32$ ;  $\rho = 0.34$ ) [46]. For each of these firing rates, the simulation was repeated 100 times, producing 100 simulated trials of the fMRI BOLD time series with temporal resolution 1 ms for a total of 25 seconds. We downsampled the time series to a sampling rate of 1 Hz to reflect the temporal resolution of real BOLD signals. We then empirically estimated the signal and noise as defined above. In particular, we investigated different subspaces spanned by  $\mathbf{U}_{t-1}$ , including (1) linear bases with different time lags; (2) second-order polynomial bases with different time lags; and (3) sixth-order Fourier bases with different time lags, to ensure that the observed signal-dependent noise phenomenon does not depend on the selection of

the projection space. Any non-random relationship between the empirically estimated signal  $\mathbf{E}(x_t|\mathbf{U}_{t-1})$  and the empirically estimated variance of the noise  $\text{var}(x_t|\mathbf{U}_{t-1})$  indicates the presence of signal-dependent noise. In particular, we tested if there exists a significant correlation between  $[\mathbf{E}(x_t|\mathbf{U}_{t-1})]^2$  and  $\text{var}(x_t|\mathbf{U}_{t-1})$ .

### fMRI experiment

The fMRI dataset is the same as that obtained and used in previous investigations [17,47,48]. We describe key imaging acquisition, preprocessing and psychophysiological interaction (PPI) analyses for completeness. We refer the readers to previous publications for the full details.

**Participants and ethics statement.** Twelve healthy volunteers (6 male and 6 female, age range 21–35) participated in the study. Ethical approval (Central Oxford Research Ethics Committee) and written informed consent from all subjects were obtained before the experiment. The subjects had not eaten for three hours before the investigation.

**Experimental design.** We used the identical taste stimulus, 0.1 M monosodium glutamate (MSG) with 0.005 M inosine monophosphate (see [49]), referred to throughout this paper for brevity as monosodium glutamate, in two different types of trial. A trial started 5 seconds before the taste delivery with the visual attentional instruction either “Remember and Rate Pleasantness” or “Remember and Rate Intensity”, which was shown until the end of the taste period. The 0.75 ml taste stimulus was delivered at  $t = 5$  s. The taste period was from  $t = 5$  s until  $t = 14$  s, and in this period a red cross was also present indicating that swallowing should not occur. The differences between the activations in this period were a measure of the effects of the top-down selective attention instructions while the taste was being delivered. (We note that in order to utilize top-down attention, one needs to hold the object of attention in mind, in this case pleasantness or intensity. This requires a short-term memory. Short-term memory is thus a *sine qua non* of selective attention [50,51], and it is the source of this top-down bias from a short-term memory system in which we are interested in this investigation.) After the end of the taste period, the visual instruction and red cross were turned off, and a green cross was shown cueing the subject to swallow. After 2 s a tasteless rinse was delivered with a red cross, and the rinse period was from  $t = 16$  s until  $t = 23$  s, when the green cross appeared to cue a swallow. After this the rating of pleasantness or intensity was made using button-press operated visual analog, rating scales ranging continuously from +2 (very pleasant) to −2 (very unpleasant) for pleasantness, and 4 (intense) to 0 (very weak) for intensity as described previously [52]. These two trial types were interspersed in random permuted sequence with other trials that were part of a different investigation, and each was presented 9 times. As different trial types were being run in the scanner at the same time, and included different stimuli, and no instructions were given about the number of stimuli being used, or that the stimuli were the same on the “Remember and Rate Intensity” and “Remember and Rate Pleasantness” trials, the participants simply had to concentrate on following the instructions about what aspect of the taste stimulus, intensity or pleasantness, had to be rated on that trial. The protocol and design are described in [17], and have been used successfully in previous studies to investigate taste cortical areas [22,53–55].

**fMRI data acquisition.** Images were acquired with a 3.0-T VARIAN/SIEMENS whole-body scanner at the Centre for Functional Magnetic Resonance Imaging at Oxford (FMRIB), where 27  $T_2^*$  weighted EPI coronal slices with in-plane resolution of  $3 \times 3$  mm and between plane spacing of 4 mm were acquired

every 2 seconds ( $TR = 2$  s). We used the techniques that we have developed over a number of years [49,53], and as described in detail by [56] we carefully selected the imaging parameters in order to minimize susceptibility and distortion artefact in the orbitofrontal cortex. The relevant factors include imaging in the coronal plane, minimizing voxel size in the plane of the imaging, as high a gradient switching frequency as possible (960 Hz), a short echo time of 28 ms, and local shimming for the inferior frontal area. The matrix size was  $64 \times 64$  and the field of view was  $192 \times 192$  mm. Continuous coverage was obtained from +62 (A/P) to −46 (A/P). A whole brain  $T_2^*$  weighted EPI volume of the above dimensions, and an anatomical  $T_1$  volume with coronal plane slice thickness 3 mm and in-plane resolution of  $1 \times 1$  mm were also acquired.

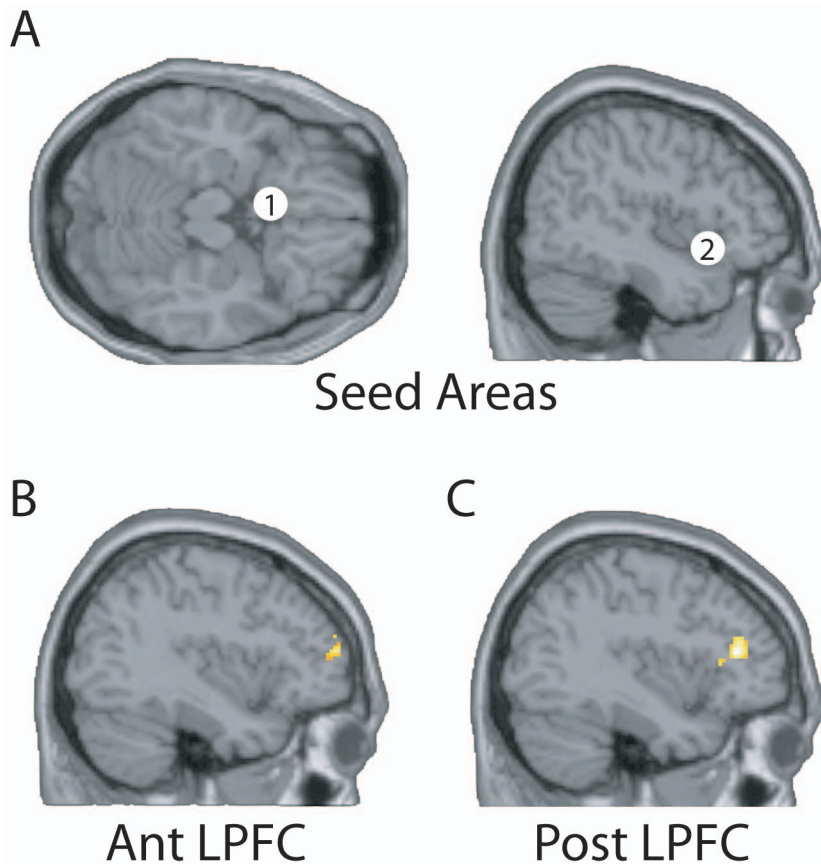
**fMRI data preprocessing.** The imaging data were analyzed using SPM5 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London. <http://www.fil.ion.ucl.ac.uk/spm/>). Preprocessing of the data used SPM5 realignment, reslicing with sinc interpolation, normalization to the Montreal Neurological Institute (MNI) coordinate system [57], and spatial smoothing with a 6 mm full width at half maximum (FWHM) isotropic Gaussian kernel. Time series non-sphericity at each voxel was estimated and corrected for [58], and a high-pass filter with a cut-off period of 128 seconds was applied.

**fMRI data analysis.** To investigate task dependent activations of brain areas during the taste period, a Finite Impulse Response (FIR) analysis was performed as implemented in SPM, in order to make no assumption about the time course based on the temporal filtering property of the haemodynamic response function [59,60]. The *a priori* defined areas of interest (ROI) for which we reported results [17] included brain areas where activations to taste stimuli have been found in previous studies including the medial and lateral orbitofrontal cortex, the pregenual part of the cingulate cortex, and the taste and oral somatosensory parts of the insular cortex [22,49,53–55,61,62]; and areas of the lateral prefrontal cortex where activations related to task set, attentional instructions, and remembering rules that guide task performance have been found, including specifically parts of the middle and inferior frontal gyrus [63–70]. A contrast of trials where attention was being paid to taste pleasantness with trials where attention was to intensity revealed significant effects in the orbitofrontal cortex [−6, 14, −20]. The reverse contrast of trials where attention was to intensity vs trials where attention was to pleasantness revealed significant effects in the right anterior insular taste cortex [42, 18, −14] [17].

We then performed PPI analyses [71,72], using the above two brain areas as seed regions, to investigate task-dependent functional connectivity of these areas with other brain areas, that might provide the source of the top-down modulation [47]. We identified an anterior lateral prefrontal cortex (AntLPFC) region at  $Y = 53$  mm in which the correlation with activity in the orbitofrontal cortex (OFC) seed region was greater when attention was to pleasantness than to intensity [47]. Conversely, in a more posterior region of lateral prefrontal cortex (PostLPFC) at  $Y = 34$  mm the correlation with activity in the anterior insula (AntINS) seed region was greater when attention was to intensity than to pleasantness [47]. The locations of the seed regions and the identified foci in AntLPFC and PostLPFC are shown in Figure 1.

**Empirical analysis of BOLD signals.** An empirical analysis was performed to provide evidence on whether there is signal-dependent noise in fMRI BOLD time series. Again, we empirically estimated the signal and noise by projecting the current state of the observed fMRI BOLD time series onto a subspace spanned by its past history. We investigated subspaces spanned by different sets of





**Figure 1. Results of the PPI analysis.** A. The seed areas for the PPI analysis in the orbitofrontal cortex (1)  $[-6, 14, -20]$ , and insular taste cortex (2)  $[42, 18, -14]$ . B. The region of the anterior lateral prefrontal cortex (AntLPFC)  $[-40, 54, 14]$  identified by PPI analysis as correlated with the orbitofrontal cortex seed area when attention was to pleasantness ( $p < 0.029$ ). C. The region of the posterior lateral prefrontal cortex  $[-38, 34, 14]$  identified by PPI analysis as correlated with the insular taste cortex seed area when attention was to intensity ( $p < 0.049$ ). The full details of the PPI analysis are provided in [47].

doi:10.1371/journal.pcbi.1003265.g001

basis functions including (1) linear bases with different time lags; (2) second-order polynomial bases with different time lags; and (3) sixth-order Fourier bases with different time lags, to ensure that the observed signal-dependent noise phenomenon does not depend on the selection of the projection space. Any non-random relationship between the estimated signal  $\mathbf{E}(x_t|\mathbf{U}_{t-1})$  and the estimated variance of the noise  $\text{var}(x_t|\mathbf{U}_{t-1})$  indicates the presence of signal-dependent noise. In particular, we tested if there exists a significant correlation between  $[\mathbf{E}(x_t|\mathbf{U}_{t-1})]^2$  and  $\text{var}(x_t|\mathbf{U}_{t-1})$ . We also investigated the correlation between the empirically estimated signal,  $[\mathbf{E}(x_t|\mathbf{U}_{t-1})]^2$ , and the model estimate of the signal,  $\hat{a}x_{t-1}$ ; and the correlation between the empirically estimated variance of the noise,  $\text{var}(x_t|\mathbf{U}_{t-1})$ , and the variance of the noise estimated by the model,  $\hat{c}^2 + \hat{b}^2x_{t-1}^2$ , where  $\hat{a}$ ,  $\hat{b}$  and  $\hat{c}$  are estimates of the model parameters, to test whether there is good concordance between the model and the empirical data analysis with respect to the signal and noise.

**Granger causal analysis of fMRI BOLD signals.** The PPI analyses described above do not show the directionality of the influences, as they are based on correlations, and for that reason we applied Granger causal analysis to each pair of the four brain areas (OFC, AntINS, AntLPFC and PostLPFC) [48]. We extracted the mean BOLD signals from 33 voxels within a sphere of radius 2 voxels centered at the seed voxels in OFC and AntINS, and the peak voxels identified with the largest PPI effect in

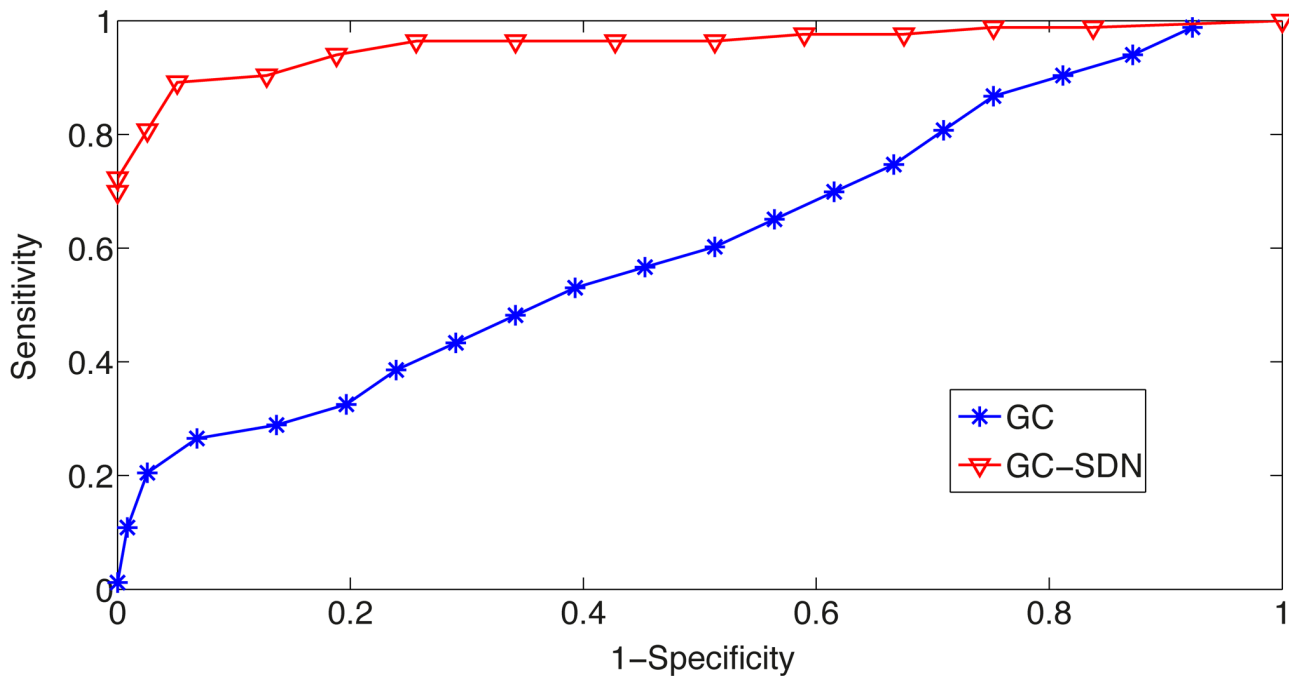
AntLPFC and PostLPFC, for Granger causal analysis. For each of the two experimental conditions (attention to intensity vs attention to pleasantness), the time series for a single subject consisted of 9 trials, each with 18 BOLD signal data points (2 s apart), starting on each trial at the onset of the instruction to pay attention to the pleasantness or to the intensity of the taste. Each trial was denoised by wavelet using a Matlab routine, **mswden**, with the Daubechies 2 (db2) wavelet, and threshold options **sqtwoolog** (universal threshold at  $\sqrt{2 \log(\cdot)}$ ) and **sln** (rescaling using a single estimation of level noise, based on first level coefficients). Each trial was also detrended and centered to zero mean before causal analyses. For each experimental condition and each pair of the four brain areas, we pooled data from all subjects ( $12 \times 9$  trials) to fit the signal-dependent noise model, i.e., we treated the 108 trials as repeated realizations from a common underlying model. We detected unidirectional causal influences as well as significant difference of the causalities in opposite directions [2] to identify the dominant causal influences in a particular direction [48]. We also applied classical GC to the same data set as a comparison.

## Results

### Simulation results

Figure 2 shows a comparison of performance for the classical Granger causal model and the Granger causal model with signal-dependent noise by ROC (receiver operating characteristic)





**Figure 2. Comparison by simulations of the performance of the classical Granger causal analysis and the Granger causality with signal-dependent noise analysis by ROC (receiver operating characteristic) analysis.** The sensitivity of the methods is plotted against  $1 - \text{specificity}$  for different  $p$ -value thresholds. The sensitivity is defined as the proportion of actual causal influences that are correctly identified. The specificity measures the proportion of non-causal influences that are correctly identified. By setting different  $p$ -value thresholds for causality, each method gives different sensitivity and specificity. Therefore, the best model is expected to have its performance ROC curve go through the upper left corner, while a random classification algorithm has its performance curve as a diagonal line. The signal-dependent noise model outperforms the classical Granger causal model substantially and consistently.  
doi:10.1371/journal.pcbi.1003265.g002

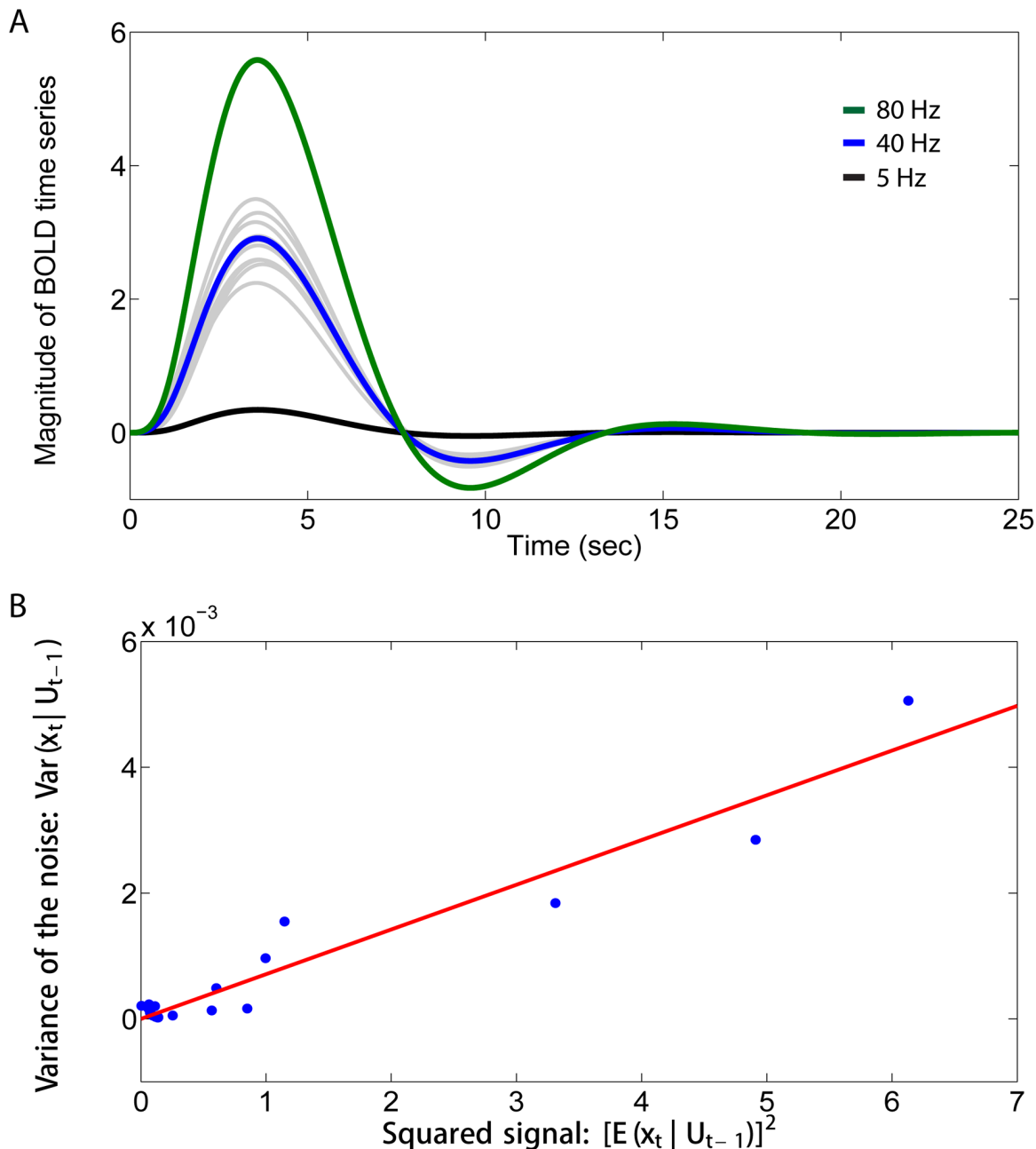
analysis. Clearly, classical GC cannot capture the causal influences well in the presence of signal-dependent noise, while the signal-dependent noise Granger causal model substantially outperforms the classical GC model, and shows a good sensitivity and specificity.

Figure 3A shows the mean BOLD signals calculated across the trials for the three firing rates before downsampling. As expected, higher firing rates evoked larger mean modelled haemodynamic responses. However, the variability of the modelled BOLD response was considerable, as illustrated for the mean firing rate of 40 Hz in Figure 3A for 10 randomly selected trials. Figure 3B shows the relation between the empirically estimated variance of the noise and the squared empirically estimated signal at different time points within a trial, using the projection space spanned by the second-order polynomial basis, i.e.,  $\text{span}\{x_{t-1}, x_{t-1}^2, x_{t-2}, x_{t-2}^2\}$ . This shows that the variance of the noise at any point in the time course of a trial is approximately linearly related to the square of the signal. We obtained consistent results using different projection spaces. Consistent results with those just described can also be obtained with a simpler model in which the spike trains are convolved with the canonical haemodynamic response function to generate the BOLD signal, as described previously [73,74]. This simple generative model of BOLD signals thus confirms that Poisson spike trains could produce fMRI BOLD time series in which the variance of the noise across the time course of a trial would be linearly related to the squared signal. We show below that this is also exactly what was found empirically in the fMRI data.

#### Empirical evidence for signal-dependent noise in BOLD signals

Figure 4 shows the empirically estimated variance of the noise in the fMRI BOLD time series obtained in this investigation as a

function of the squared empirically estimated signal at each time point within a trial, using the projection space spanned by the second-order polynomial basis, i.e.,  $\text{span}\{x_{t-1}, x_{t-1}^2, x_{t-2}, x_{t-2}^2\}$ . Significant correlations are observed for both experimental conditions by pooling data from the four brain regions (attention to intensity,  $C=0.56$ ,  $p=4.17 \times 10^{-6}$ , attention to pleasant,  $C=0.42$ ,  $p=7.75 \times 10^{-4}$ ), which clearly indicates the presence of signal-dependent noise in the fMRI BOLD time series. In particular, the results shown in Figure 4 show that the variance of the noise in BOLD time series is approximately linearly related to the squared signal. A similar effect was also found for each brain region when analyzed separately. The results were consistent using different projection spaces. In particular, we observed significant correlations when the project space was spanned by (1) linear bases up to 9 time lags; (2) second-order polynomial bases up to 6 time lags; and (3) sixth-order Fourier bases up to 2 time lags. These results provide strong evidence for the presence of signal-dependent noise in fMRI BOLD time series. Moreover, when fitting our signal-dependent noise model to the real data, we observed excellent concordance and significant correlation between the empirically estimated signal,  $[E(x_t|U_{t-1})]^2$ , and the model estimate of the signal,  $\hat{a}x_{t-1}$  (attention to intensity,  $C=0.35$ ,  $p=4.00 \times 10^{-3}$ , attention to pleasantness,  $C=0.78$ ,  $p=2.88 \times 10^{-14}$ ), and between the empirically estimated variance of the noise,  $\text{var}(x_t|U_{t-1})$ , and the variance of the noise estimated by the model,  $\hat{c}^2 + \hat{b}^2 x_{t-1}^2$  (attention to intensity,  $C=0.56$ ,  $p=1.17 \times 10^{-6}$ , attention to pleasantness,  $C=0.67$ ,  $p=1.90 \times 10^{-9}$ ). The results were also consistent using different projection spaces. This indicates that the AR-BEKK model is a good



**Figure 3. An illustrative model of signal-dependent noise in BOLD signals.** A. The mean BOLD signals for the different time points within a trial (calculated across the trials) for three firing rates, with mean rates of 5, 40 and 80 spikes/sec. 10 randomly selected trials of the BOLD signals with the input firing rate of 40 Hz are also shown (gray). B. The empirically estimated variance of the noise in the simulated BOLD time series plotted against the squared empirically estimated signal using the projection space spanned by the second-order polynomial basis, with the input firing rate of 40 Hz. The relation is approximately linear. doi:10.1371/journal.pcbi.1003265.g003

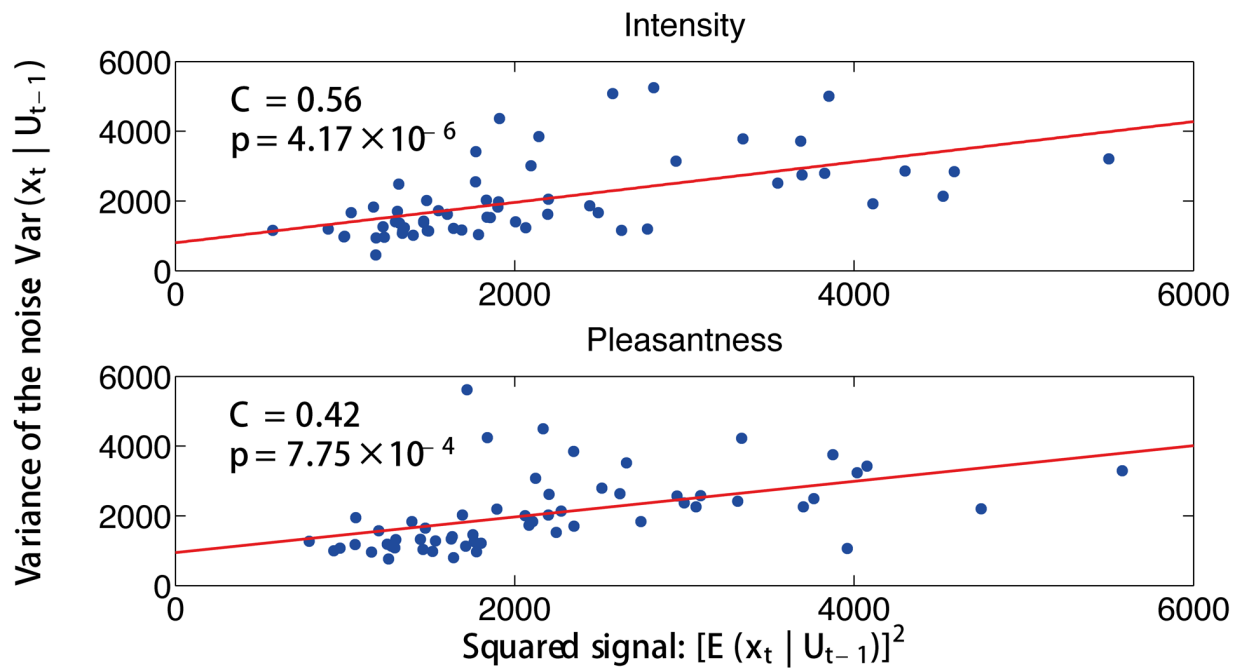
description of the data and captures a large portion of the variance in the empirical signal and noise.

#### fmRI data investigation

Table 1 shows the causal influences between the four brain areas (OFC, AntINS, AntLPFC, PostLPFC) detected by the Granger causality with signal-dependent noise analysis. First, we consider attention to intensity. There are significant (top-down) causal influences from both the AntLPFC and PostLPFC to the insular taste cortex (AntINS). Second, we consider attention to pleasantness.

There are significant (top-down) causal influences from both the AntLPFC and PostLPFC to the OFC, and a significant effect from the OFC to the antLPFC. There is also a (top-down) effect of the PostLPFC on the taste insula (AntINS). Very interestingly too, during attention to pleasantness, there is increased effective connectivity from the insular taste cortex to the OFC, suggesting that information is routed especially to the OFC during attention to pleasantness.

For comparison, Table 2 shows the causal influences between the four brain areas detected by the classical Granger causal model. Only one effective connectivity influence (PostLPFC to



**Figure 4. Empirical evidence for signal-dependent noise in BOLD signals.** For each of the four brain areas and each subject, the empirically estimated variance of noise in the observed fMRI BOLD time series is plotted against the squared empirically estimated signal for each time point within a trial using the projection space spanned by the second-order polynomial basis.  $U_t$  is defined in the text, and reflects the past history of times series  $x_t$ . Significant correlations are observed for both experimental conditions (attention to intensity,  $C=0.56$ ,  $p=4.17 \times 10^{-6}$ , and attention to pleasantness,  $C=0.42$ ,  $p=7.75 \times 10^{-4}$ ). doi:10.1371/journal.pcbi.1003265.g004

AntLPFC, when paying attention to intensity) was identified as significant. The greater power of the signal-dependent noise model can be clearly observed.

Table 3 shows the difference of the causalities in opposite directions by the Granger causality with signal-dependent noise analysis. In the pleasantness condition, consistent with the hypothesis that the lateral prefrontal cortex is the source of the top-down modulation of activations in the OFC, there are significantly stronger effects from both the AntLPFC and the PostLPFC to the OFC than vice versa. It is also of interest that in the pleasantness condition, a significantly stronger forward influence was detected from the AntINS to the OFC. Only one significant difference was detected for the intensity condition, that is the effect from the PostLPFC to AntINS is greater than in the reverse direction. This is consistent with the hypothesis that the major top-down effect on the taste insula during attention to intensity is from the PostLPFC. The bi-directional interaction in the pleasantness condition between the AntLPFC and OFC (Table 1) may be interpreted in the context that there is a significant difference of the causality with AntLPFC to OFC greater than OFC to AntLPFC, thus indicating a stronger influence of AntLPFC on OFC than vice versa (Table 3).

These analyses provide evidence for the effective connectivities in the attention to intensity and pleasantness conditions that are summarized in Figure 5.

## Discussion

In this paper, we for the first time provide empirical evidence for signal-dependent noise in fMRI BOLD signals in several cortical areas, such as the insular, orbitofrontal, and lateral prefrontal cortical areas. We then developed a Granger causal model with

signal-dependent noise that can appropriately model BOLD signals and detect causal influences in both mean and variance. By simulation studies, we showed that our Granger causality with signal-dependent noise analysis substantially outperforms classical Granger causal analysis, when signal-dependent noise is present in the time series. We applied our Granger causal model with signal-dependent noise to the data from an fMRI study to investigate the source of the top-down attentional influences on taste processing when attention was to the intensity vs the pleasantness of the taste. We found a top-down effect from the PostLPFC to the insular taste cortex during attention to intensity but not to pleasantness; and a top-down effect from the AntLPFC and PostLPFC to the OFC during attention to pleasantness but not to intensity. In addition, there was stronger forward effective connectivity from the insular taste cortex to the OFC during attention to pleasantness than during attention to intensity.

## Assessment of the measurement of Granger causality taking into account signal-dependent noise

Conditionally heteroskedastic data often show volatility clustering and outliers. In particular, the unconditional distribution of the data is leptokurtic, which means that it has more mass around zero and in the tails than the normal distribution and, hence, it can produce occasional outliers [27]. Therefore, models with time varying volatility can better capture the nature of the data, and it is expected that more reliable causal inferences can be made. Comparing to the earlier approaches of causal inferences in data with time varying volatility [75–78], including [13], the model presented in this paper that takes into account signal-dependent noise provides an accurate, efficient and unified method to detect causality in both the mean and variance. The model has a corresponding frequency domain representation [13], which may

**Table 1.** Causality results by Granger causality with signal-dependent noise analysis.

Intensity				
	OFC	AntINS	AntLPFC	PostLPFC
OFC	–	4.51 (0.10)	0.96 (0.62)	0.68 (0.71)
AntINS	2.15 (0.34)	–	<b>28.57</b> ( $< 10^{-4}$ )	13.59 ( $1.1 \times 10^{-3}$ )
AntLPFC	0.89 (0.64)	<b>28.44</b> ( $< 10^{-4}$ )	–	5.17 (0.08)
PostLPFC	0.62 (0.73)	<b>29.62</b> ( $< 10^{-4}$ )	4.22 (0.12)	–
Pleasantness				
	OFC	AntINS	AntLPFC	PostLPFC
OFC	–	16.92 ( $2 \times 10^{-4}$ )	<b>44.94</b> ( $< 10^{-4}$ )	7.12 (0.03)
AntINS	<b>94.79</b> ( $< 10^{-4}$ )	–	2.09 (0.35)	17.56 ( $2 \times 10^{-4}$ )
AntLPFC	<b>95.08</b> ( $< 10^{-4}$ )	6.72 (0.035)	–	6.93 (0.03)*
PostLPFC	<b>89.16</b> ( $< 10^{-4}$ )	<b>22.81</b> ( $< 10^{-4}$ )	5.70 (0.06)	–

The causal influence is from row to column. The causality is given for each direction, and the corresponding  $p$ -value is presented in brackets. If the uncorrected  $p$ -value is less than  $10^{-4}$  (surviving Bonferroni correction), the causal influence is identified as significant and indicated in bold in the table. (\*: The active-set algorithm was used.)  
doi:10.1371/journal.pcbi.1003265.t001

further shed light on frequency-specific interactions. The model described here applies when the variance of the noise is proportional to the square of the signal, which is what we observed from real fMRI BOLD time series, but could in principle be extended to deal with other cases. There are alternative measures of Granger-type causality such as partial directed coherence (PDC) [79], relative power contribution (RPC) [80] and directed transfer function (DTF) [81] that do not explicitly use the noise covariance function to define causality but are based on the transfer function or model coefficients. However, because they are typically formulated under the simple AR model, all these methods are unable to capture causal influences in the second order statistics such as signal-dependent noise.

It is not easy to tease apart ‘signal’ and ‘noise’ from an observed time series. In this paper, we define ‘signal’ as the part of the observations that can be well predicted from the past history of the time series, and ‘noise’ as what is completely unpredictable and produces the variation across realizations. We therefore empirically

estimate the signal by projecting the current state of the time series onto the subspace spanned by its past history. In practice, if the projection space is not constructed appropriately, part of the signal that does not lie in the projection space may migrate to the residual process and produce artificial signal-dependent noise phenomena. In this case, expanding the projection space, i.e., using a more complex model to describe the mean activity of the time series, may mitigate the issues of signal-dependent noise. However, if the variance of the noise is indeed dependent on the signal, simply increasing the complexity of the model in the mean structure will not remove this dependence. In the present paper, we investigated a number of projection spaces, spanned by linear or nonlinear basis functions with different time lags, and always observed signal-dependent noise. Therefore, there is strong evidence for the presence of signal-dependent noise in fMRI BOLD time series. In particular, the variance of the noise is approximately linearly related to the square of the signal. When constructing our signal-dependent noise model, we made use of this relationship and specified a linear

**Table 2.** Causality results by classical Granger causal analysis.

Intensity				
	OFC	AntINS	AntLPFC	PostLPFC
OFC	–	0.0039 (0.0061)	0.0006 (0.30)	0.0001 (0.62)
AntINS	0.0015 (0.09)	–	0.0002 (0.56)	0.0003 (0.42)
AntLPFC	0.0000 (0.99)	0.0056 (0.0009)	–	0.0043 (0.0041)
PostLPFC	0.0003 (0.43)	<b>0.0122</b> ( $< 10^{-4}$ )	0.0013 (0.12)	–
Pleasantness				
	OFC	AntINS	AntLPFC	PostLPFC
OFC	–	0.0000 (0.99)	0.0001 (0.71)	0.0000 (0.98)
AntINS	0.0001 (0.74)	–	0.0023 (0.03)	0.0018 (0.06)
AntLPFC	0.0000 (0.99)	0.0002 (0.51)	–	0.0059 (0.0007)
PostLPFC	0.0008 (0.22)	0.0034 (0.0099)	0.0002 (0.59)	–

The causal influence is from row to column. The causality is given for each direction and the corresponding  $p$ -value is presented in brackets. If the uncorrected  $p$ -value is less than  $10^{-4}$  (surviving Bonferroni correction), the causal influence is identified as significant and indicated in bold in the table.  
doi:10.1371/journal.pcbi.1003265.t002

**Table 3.** Difference of the causalities in opposite directions by the Granger causality with signal-dependent noise analysis.

Direction	Intensity	Pleasantness
PostLPFC→AntLPFC – AntLPFC→PostLPFC	–0.48	–0.614*
PostLPFC→AntINS – AntINS→PostLPFC	<b>8.02</b>	2.62
PostLPFC→OFC – OFC→PostLPFC	–0.03	<b>41.03</b>
AntLPFC→AntINS – AntINS→AntLPFC	–0.06	2.32
AntLPFC→OFC – OFC→AntLPFC	–0.03	<b>25.07</b>
AntINS→OFC – OFC→AntINS	–1.18	<b>38.93</b>

Differences with a *p*-value smaller than 0.01 (a difference measure greater than 4.61) are indicated in bold. (\*: We used the active-set algorithm for this particular link.)  
doi:10.1371/journal.pcbi.1003265.t003

function to describe the mean activity of the observations, and a quadratic function for the variance of the noise. Although the model appears to be simple, we have shown that it captures a large portion of the variance in the signal and noise in the empirical BOLD time series. Future studies will be of interest to provide more evidence on signal-dependent noise in different brain areas and different data sets, to further examine the relationship between the variance of the noise and the signal, and to develop more complex models, e.g., using nonlinear functions or kernels, accordingly.

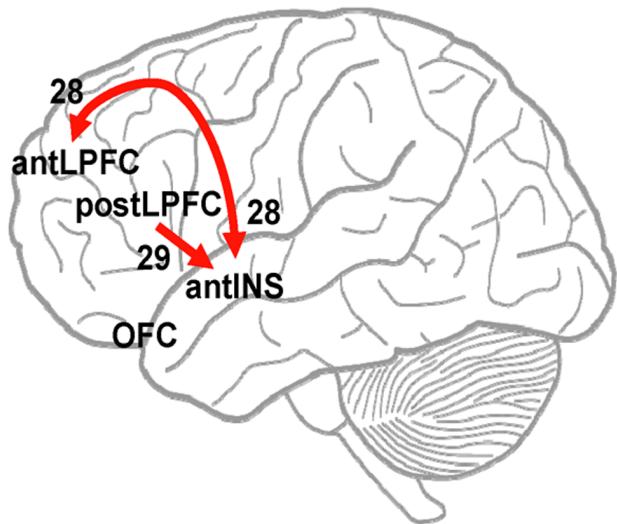
Although we only applied our model to fMRI time series, it is clear that the model can be applied to very many types of data that might exhibit signal-dependent noise, including neurophysiological data such as single or multi-neuron recordings, magnetoencephalography, local field potentials, and beyond neuroscience also to any possibly causal system where there are time series of data from two or many sources. Indeed, the significance of detecting causality from data with time varying volatility might be partly demonstrated in the 2003 Nobel Prize in Economics shared by Granger, who set up the foundation of Granger causal analysis

[6], and Engle, who invented the first changing volatility model [31].

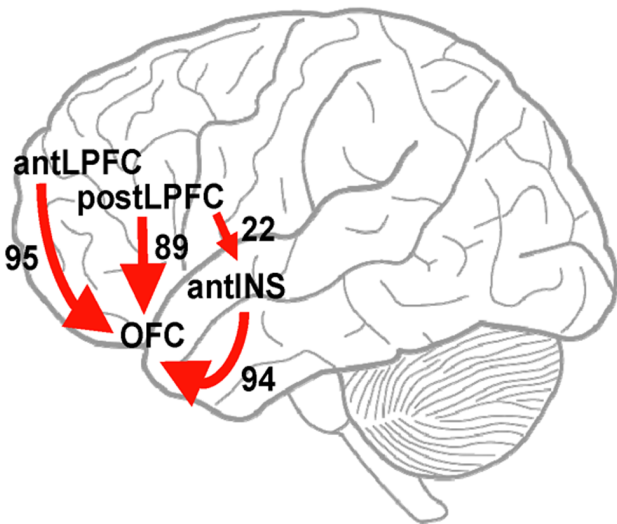
Although our initial implementation of the signal-dependent noise model appears to be successful, due to the highly nonlinear form of the log-likelihood function and optimization problem, fast and robust optimization algorithms deserve future investigation. Also, although a low-order low-dimensional AR-BEKK model is a relatively parsimonious representation of the conditional covariance structure of a process, the number of parameters still grows quickly with the dimension of the underlying system. This impedes the application of the model to a modest number of time series. Future studies are needed to find more restricted models that ensure uniqueness of the parameterization, guarantee the positive definiteness of the conditional covariance, while at the same time still produce rich dynamics.

In spite of the wide and successful applications in neurophysiological data, there is still an ongoing debate on applying GC to fMRI data [10,82–87]. Inferring causality from fMRI time series — an indirect measure of neuronal activities — imposes many more challenges than direct electrophysiological recordings. Granger causal models use the observed fMRI data as a surrogate for the underlying neuronal activity, which is a potential flaw of the method and the main controversy against the application of GC to fMRI data, since the BOLD signal is a blurred and delayed representation of the original neuronal signal, and it is now widely recognized that there is intra- and inter-subject variability of haemodynamic responses [88–91]. However, there have been a series of numerical and theoretical works showing that GC is quite robust to the difference in haemodynamic delays [92–94]. Moreover, as in [4], we calculated the cross-correlation function for each pair of time series used in our Granger causal analysis, and most of the cross-correlation peaks appeared at zero lag, indicating that differences in the regional haemodynamic responses may not be a significant factor in this study. We therefore feel that the application of Granger type causal inferences in the analysis of this particular fMRI data set is justified. However, given the complexity of the brain, much work

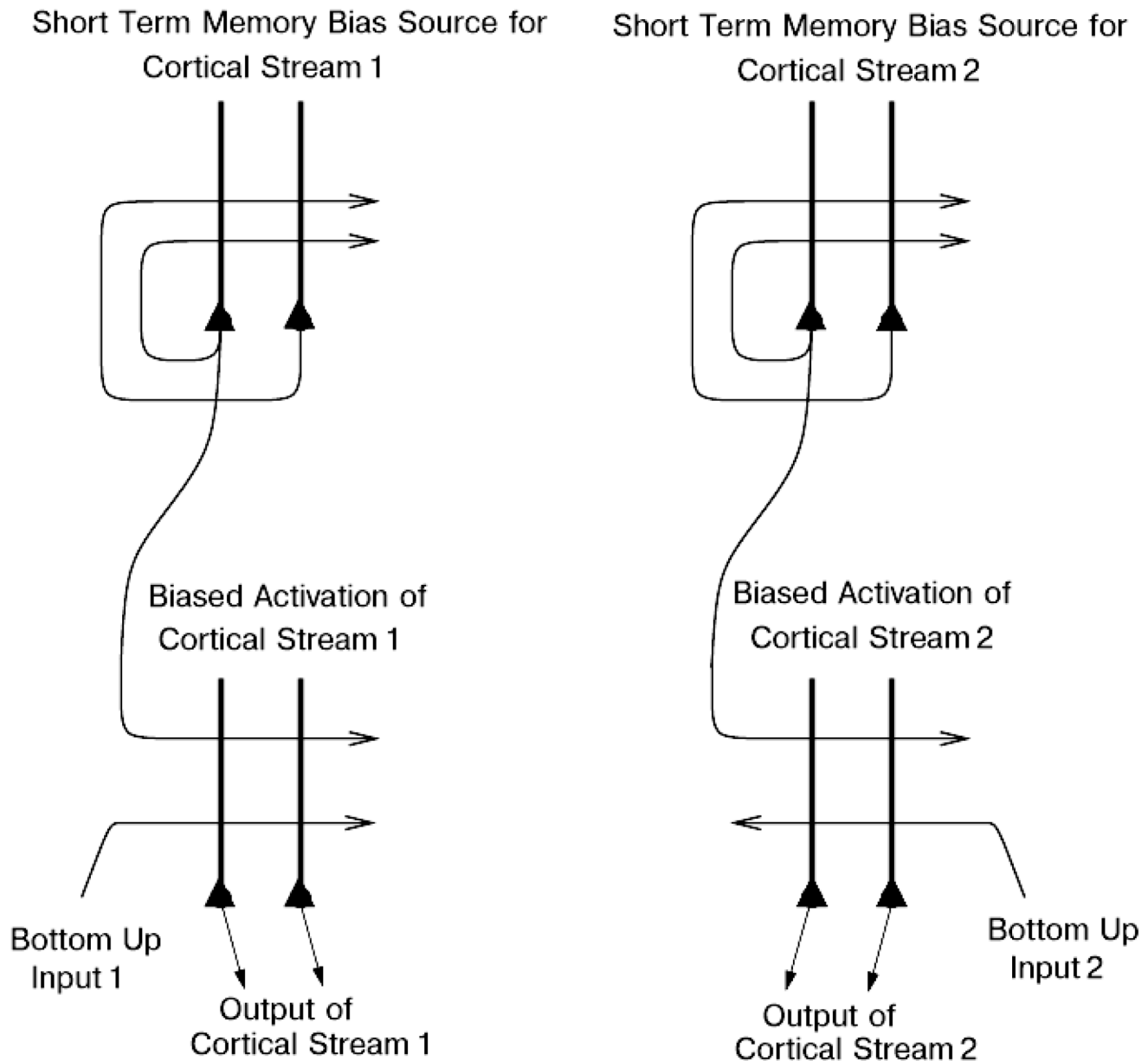
A Intensity



B Pleasantness



**Figure 5.** Neural circuits revealed by Granger causality with signal-dependent noise. A. Attention to taste intensity. B. Attention to taste pleasantness. Larger arrows represent a stronger influence. The values of significant likelihood ratio test statistics are indicated.  
doi:10.1371/journal.pcbi.1003265.g005



**Figure 6. A Biased activation theory of selective attention.** The short-term memory systems that provide the source of the top-down activations may be separate (as shown), or could be a single network with different attractor states for the different selective attention conditions. The top-down short-term memory systems hold what is being paid attention to active by continuing firing in an attractor state, and bias separately either cortical processing system 1, or cortical processing system 2. This weak top-down bias interacts with the bottom up input to the cortical stream and produces an increase of activity that can be supralinear [98]. Thus the selective activation of separate cortical processing streams can occur. In the example, stream 1 might process the affective value of a stimulus with the areas involved including the anterior lateral prefrontal cortex with a top-down influence on the orbitofrontal cortex, and stream 2 might process the intensity and physical properties of the stimulus with the areas involved including the posterior lateral prefrontal cortex with a top-down influence on the insular taste cortex. The outputs of these separate processing streams then must enter a competition system, which could be for example a cortical attractor decision-making network that makes choices between the two streams, with the choice biased by the activations in the separate streams [19].  
doi:10.1371/journal.pcbi.1003265.g006

remains to do to provide reliable and accurate causal analyses for neuroscience.

### Neural interpretation

The interpretation of the effective connectivity revealed with our signal-dependent noise model is that during attention to pleasantness, the AntLPFC and PostLPFC regions identified by PPI analysis exert a top-down control of the responsiveness of

the OFC to its taste-related inputs, and indeed to how strongly information is routed to the OFC from its preceding area, the AntINS taste cortex. In contrast, during attention to intensity, the PostLPFC identified by PPI analysis exerts a top-down control of the responsiveness of the insular taste cortex to its taste-related inputs. This interpretation is strengthened by the findings with our componential Granger causal analysis [48], which provides evidence that the top-down effects depend on



the level of activity in the areas on which there is a top-down effect.

The way that we think of top-down biased competition as operating normally in, for example, visual selective attention [95] is that within an area, e.g. a cortical region, some neurons receive a weak top-down input that increases their response to the bottom-up stimuli [95], potentially supra-linearly if the bottom-up stimuli are weak [50,51,63]. The enhanced firing of the biased neurons then, via the local inhibitory neurons, inhibits the other neurons in the local area from responding to the bottom-up stimuli. This is a local mechanism, in that the inhibition in the neocortex is primarily local, being implemented by cortical inhibitory neurons that typically have inputs and outputs over no more than a few mm [50,51,96]. This model of biased competition is illustrated in [47]. That locally implemented biased competition situation may not apply in the present case, where we have facilitation of processing in a whole cortical area (e.g. orbitofrontal cortex) or even cortical processing stream (e.g. the linked orbitofrontal and pregenual cingulate cortex [47]) in which the activity of taste neurons may reflect pleasantness and not intensity. So the

attentional effect might more accurately be described in the present case as biased activation, without local competition being part of the effect. This biased activation theory and model of attention, illustrated in Figure 6, is a rather different way to implement attention in the brain than biased competition, and each mechanism may apply in different cases, or both mechanisms in some cases [19,47,97].

## Acknowledgments

We are grateful to Dr. Wanlu Deng for many helpful discussions. We have made some of the software used in the analysis of Granger causality with signal-dependent noise available at <http://www.dcs.warwick.ac.uk/~feng/causality.html>.

## Author Contributions

Conceived and designed the experiments: QL TG FG JF ETR. Performed the experiments: QL TG FG ETR. Analyzed the data: QL TG FG ETR. Contributed reagents/materials/analysis tools: QL TG FG JF ETR. Wrote the paper: QL TG FG JF ETR.

## References

- Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging* 21: 1251–1261.
- Roebroeck A, Formisano E, Goebel R, et al. (2005) Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25: 230–242.
- Hwang K, Velanova K, Luna B (2010) Strengthening of top-down frontal cognitive control networks underlying the development of inhibitory control: a functional magnetic resonance imaging effective connectivity study. *Journal of Neuroscience* 30: 15535–15545.
- Wen X, Yao L, Liu Y, Ding M (2012) Causal interactions in attention networks predict behavioral performance. *Journal of Neuroscience* 32: 1284–1292.
- Wiener N (1956) The theory of prediction. *Modern mathematics for engineers*. New York: McGraw-Hill: 165–190.
- Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* 37: 424–438.
- Ding M, Chen Y, Bressler S (2006) Granger causality: Basic theory and application to neuroscience. In: Schelter B, Winterhalder M, Timmer J, editors. *Handbook of Time Series Analysis*. Weinheim: Wiley-VCH.
- Friston K, Moran R, Seth A (2012) Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology* 23: 1–7.
- Ge T, Kendrick KM, Feng J (2009) A novel extended Granger causal model approach demonstrates brain hemispheric differences during face recognition learning. *PLoS Computational Biology* 5: e1000570.
- David O, Guillemain I, Saitlet S, Rey S, Deransart C, et al. (2008) Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biology* 6: e315.
- Gerstein GL, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron. *Biophysical Journal* 4: 41–68.
- McAdams CJ, Maunsell JHR (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23: 765–773.
- Luo Q, Ge T, Feng J (2011) Granger causality with signal-dependent noise. *NeuroImage* 57: 1422–1429.
- Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. *Nature* 394: 780–784.
- Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation* 17: 1084–1108.
- Selen LPJ, Franklin DW, Wolpert DM (2009) Impedance control reduces instability that arises from motor noise. *The Journal of Neuroscience* 29: 12606–12616.
- Grabenhorst F, Rolls ET (2008) Selective attention to affective value alters how the brain processes taste stimuli. *European Journal of Neuroscience* 27: 723–729.
- Rolls ET, Grabenhorst F, Margot C, da Silva M, Velasco M (2008) Selective attention to affective value alters how the brain processes olfactory stimuli. *Journal of Cognitive Neuroscience* 20: 1815–1826.
- Grabenhorst F, Rolls ET (2011) Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences* 15: 56–67.
- de Araujo IET, Kringelbach ML, Rolls ET, McGlone F (2003) Human cortical responses to water in the mouth, and the effects of thirst. *Journal of Neurophysiology* 90: 1865–1876.
- Kringelbach ML, O'Doherty J, Rolls ET, Andrews C (2003) Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cerebral Cortex* 13: 1064–1071.
- Grabenhorst F, Rolls ET, Bilderbeck A (2008) How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cerebral Cortex* 18: 1549–1559.
- Grabenhorst F, Rolls ET, Parris BA, d'Souza AA (2010) How the brain represents the reward value of fat in the mouth. *Cerebral Cortex* 20: 1082–1091.
- Rolls ET, Grabenhorst F (2008) The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in Neurobiology* 86: 216–244.
- Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, et al. (2003) Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39: 701–711.
- Haase L, Cerf-Ducastel B, Murphy C (2009) Cortical activation in response to pure taste stimuli during the physiological states of hunger and satiety. *NeuroImage* 44: 1008–1021.
- Lütkepohl H (2005) New introduction to multiple time series analysis. Cambridge Univ Press.
- Geweke JF (1982) Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association* 77: 304–313.
- Geweke JF (1984) Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* 79: 907–915.
- Barnett L, Bossomaier T (2012) Transfer entropy as a log-likelihood ratio. *Physical Review Letters* 109: 138105.
- Engle R (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society* : 987–1007.
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–327.
- Taylor S (1986) Modelling financial time series. *World Scientific* 113: 266.
- Engle R, Granger CWJ, Kraft D (1984) Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control* 8: 151–165.
- Bollerslev T, Engle RF, Wooldridge J (1988) A capital asset pricing model with time-varying covariances. *The Journal of Political Economy* : 116–131.
- Diebold FX, Nerlove M (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4: 1–21.
- Baba Y, Engle RF, Kraft DF, Kroner KF (1991) Multivariate simultaneous generalized ARCH. Discussion paper, University of California, San Diego, Department of Economics 11: 122–150.
- Engle RF, Kroner KF (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11: 122–150.
- Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience* 28: 10056–10061.
- Hamilton JP, Chen G, Thomason M, Schwartz M, Gotlib I (2010) Investigating neural primacy in major depressive disorder: multivariate Granger causality analysis of resting-state fMRI time-series data. *Molecular Psychiatry* 16: 763–772.
- Knepp DL, Entwistle DR (1969) Testing significance of differences between two chi-squares. *Psychometrika* 34: 331–333.

42. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861–874.
43. Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* 12: 466–477.
44. Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine* 39: 855–864.
45. Mandeville JB, Marota JJA, Ayata C, Zaharchuk G, Moskowitz MA, et al. (1999) Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *Journal of Cerebral Blood Flow & Metabolism* 19: 679–689.
46. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *NeuroImage* 19: 1273–1302.
47. Grabenhorst F, Rolls ET (2010) Attentional modulation of affective versus sensory processing: functional connectivity and a top-down biased activation theory of selective attention. *Journal of Neurophysiology* 104: 1649–1660.
48. Ge T, Feng J, Grabenhorst F, Rolls ET (2012) Componential Granger causality, and its application to identifying the source and mechanisms of the top-down biased activation that controls attention to affective vs sensory processing. *NeuroImage* 59: 1846–1858.
49. de Araujo IET, Kringelbach M, Rolls ET, Hobden P (2003) Representation of umami taste in the human brain. *Journal of Neurophysiology* 90: 313–319.
50. Rolls ET, Deco G (2002) Computational neuroscience of vision. Oxford: Oxford University Press.
51. Rolls ET (2008) Memory, attention and decision-making. New York: Oxford University Press.
52. Rolls ET, Kringelbach ML, de Araujo IET (2003) Different representations of pleasant and unpleasant odours in the human brain. *European Journal of Neuroscience* 18: 695–703.
53. O'doherty J, Rolls ET, Francis S, Bowtell R, McGlone F (2001) Representation of pleasant and aversive taste in the human brain. *Journal of Neurophysiology* 85: 1315–1321.
54. de Araujo IET, Rolls ET, Kringelbach ML, McGlone F, Phillips N (2003) Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *European Journal of Neuroscience* 18: 2059–2068.
55. McCabe C, Rolls ET (2007) Umami: a delicious flavor formed by convergence of taste and olfactory pathways in the human brain. *European Journal of Neuroscience* 25: 1855–1864.
56. Wilson JL, Jenkinson M, de Araujo I, Kringelbach ML, Rolls ET, et al. (2002) Fast, fully automated global and local magnetic field optimization for fMRI of the human brain. *NeuroImage* 17: 967–976.
57. Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography* 18: 192.
58. Friston KJ, Glaser DE, Henson RNA, Kiebel S, Phillips C, et al. (2002) Classical and Bayesian inference in neuroimaging: applications. *NeuroImage* 16: 484–512.
59. Gottfried JA, Winston JS, Dolan RJ (2006) Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron* 49: 467–479.
60. Yacubian J, Gläscher J, Schroeder K, Sommer T, Braus D, et al. (2006) Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *The Journal of Neuroscience* 26: 9530–9537.
61. Schoenfeld MA, Neuer G, Tempelmann C, Schüßler K, Noesselt T, et al. (2004) Functional magnetic resonance tomography correlates of taste perception in the human primary taste cortex. *Neuroscience* 127: 347–353.
62. Nitschke JB, Dixon G, Sarinopoulos I, Short SJ, Cohen JD, et al. (2006) Altering expectancy dampens neural response to aversive taste in primary taste cortex. *Nature Neuroscience* 9: 435–442.
63. Deco G, Rolls E (2005) Attention, short-term memory, and action selection: a unifying theory. *Progress in Neurobiology* 76: 236–256.
64. Sakai K, Passingham RE (2006) Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. *The Journal of Neuroscience* 26: 1211–1218.
65. Sakai K, Passingham RE (2002) Prefrontal interactions reflect future task operations. *Nature Neuroscience* 6: 75–81.
66. Veldhuizen MG, Bender G, Constable RT, Small DM (2007) Trying to detect taste in a tasteless solution: modulation of early gustatory cortex by attention to taste. *Chemical Senses* 32: 569–581.
67. Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research* 49: 1154–1165.
68. Rossi AF, Pessoa L, Desimone R, Ungerleider LG (2009) The prefrontal cortex and the executive control of attention. *Experimental Brain Research* 192: 489–497.
69. Kounchev F, Charron S, Koechlin E (2009) Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience* 12: 939–945.
70. Bengtsson SL, Haynes JD, Sakai K, Buckley MJ, Passingham RE (2009) The representation of abstract task rules in the human prefrontal cortex. *Cerebral Cortex* 19: 1929–1936.
71. Friston KJ, Buechel C, Fink GR, Morris J, Rolls ET, et al. (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6: 218–229.
72. Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiological interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage* 19: 200–207.
73. Rolls ET, Grabenhorst F, Deco G (2010) Choice, difficulty, and confidence in the brain. *NeuroImage* 53: 694–706.
74. Rolls ET, Grabenhorst F, Deco G (2010) Decision-making, errors, and confidence in the brain. *Journal of Neurophysiology* 104: 2359–2374.
75. Cheung YW, Ng LK (1996) A causality-in-variance test and its application to financial market prices. *Journal of Econometrics* 72: 33–48.
76. Hong YM (2001) A test for volatility spillover with application to exchange rates. *Journal of Econometrics* 103: 183–224.
77. Pantelidis T, Pittis N (2004) Testing for Granger causality in variance in the presence of causality in mean. *Economics Letters* 85: 201–207.
78. Hafner CM, Herwartz H (2008) Testing for causality in variance using multivariate GARCH models. *Annales d'Economie et de Statistique* 89: 215–241.
79. Baccalà L, Sameshima K (2001) Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics* 84: 463–474.
80. Yamashita O, Sadato N, Okada T, Ozaki T (2005) Evaluating frequency-wise directed connectivity of BOLD signals applying relative power contribution with the linear multivariate time-series models. *NeuroImage* 25: 478–490.
81. Kamiński M, Ding M, Truccolo W, Bressler S (2001) Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics* 85: 145–157.
82. Friston KJ (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biology* 7: e1000033.
83. Roebroeck A, Formisano E, Goebel R (2011) The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58: 296–302.
84. Friston KJ (2011) Dynamic causal modeling and Granger causality. Comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58: 303–305.
85. David O (2011) fMRI connectivity, meaning and empiricism. Comments on: Roebroeck et al. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58: 306–309.
86. Roebroeck A, Formisano E, Goebel R (2011) Reply to Friston and David: After comments on: The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *NeuroImage* 58: 310–311.
87. Valdes-Sosa PA, Roebroeck A, Daunizeau J, Friston KJ (2011) Effective connectivity: Influence, causality and biophysical modeling. *NeuroImage* 58: 339–361.
88. Rajapakse JC, Kruggel F, Maisog JM, Von Cramon DY (1998) Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping* 6: 283–300.
89. Aguirre GK, Zarahn E, D'Esposito M (1998) The variability of human, BOLD hemodynamic responses. *NeuroImage* 8: 360–369.
90. Kruggel F, Von Cramon DY (1999) Temporal properties of the hemodynamic response in functional MRI. *Human Brain Mapping* 8: 259–271.
91. Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage* 21: 1639–1651.
92. Deshpande G, Sathian K, Hu X (2010) Effect of hemodynamic variability on Granger causality analysis of fMRI. *NeuroImage* 52: 884–896.
93. Schippers MB, Renken R, Keysers C (2011) The effect of intra-and inter-subject variability of hemodynamic responses on group level Granger causality analyses. *NeuroImage* 57: 22–36.
94. Barnett L, Seth AK (2011) Behaviour of Granger causality under filtering: Theoretical invariance and practical application. *Journal of Neuroscience Methods* 201: 404–419.
95. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annual review of neuroscience* 18: 193–222.
96. Shepherd G (2003) The synaptic organization of the brain. New York: Oxford University Press.
97. Rolls ET (2014) Emotion and Decision-Making. Oxford: Oxford University Press.
98. Deco G, Rolls ET (2005) Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. *Journal of Neurophysiology* 94: 295–313.